

THAT WHICH IS CLAIMED IS:

1. A computer-based method for screening a nucleic acid sequence for efficient translation in a predetermined host, said method comprising the steps of:

5 (a) providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

(b) providing a substrate nucleic acid sequence;

10 (c) determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest in the substrate;

(d) determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second substrate alignment of interest is one base downstream from said first alignment of interest;

15 (e) determining a third binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third substrate alignment of interest is two bases downstream from said first alignment of interest;

20 (f) successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

25 (g) generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

30 (h) detecting the presence or absence of a three-base binding strength periodic cycle and phase through said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle and correct phase through said substrate nucleic acid sequence indicating said substrate nucleic acid sequence is a candidate for efficient translation in said host.

2. A method according to claim 1, wherein said substrate nucleic acid sequence is heterologous to said host.

3. A method according to claim 2, wherein said substrate nucleic acid sequence is obtained from a different host than said ribosomal nucleic acid sequence.

4. A method according to claim 2, wherein said substrate nucleic acid sequence is synthetic.

5. A method according to claim 1, wherein said substrate nucleic acid sequence encodes a predetermined protein or peptide.

6. A method according to claim 1, wherein said detecting further comprises the step of determining the strength of said periodic signal.

7. A method according to claim 6, further comprising the step of:

(i) generating a quantitative indicator of translation efficiency from the strength of said periodic signal.

8. A method according to claim 6, further comprising the steps of:

(j) determining the sufficiency of said translation efficiency from said quantitative indicator; and then, in the absence of sufficient translation efficiency;

(k) replacing at least one base in said substrate nucleic acid sequence with a different base to produce a subsequent substrate nucleic acid sequence different from said previous substrate nucleic acid sequence, and encoding the same protein or peptide; and

(l) repeating steps (c) through (j) above with said subsequent substrate nucleic acid sequence.

9. A method according to claim 1, further comprising the steps of:

(j) determining the sufficiency of said translation efficiency from said quantitative indicator; and then, in the absence of sufficient translation efficiency;

(k) replacing at least one base in said substrate nucleic acid sequence with a different base to produce a subsequent substrate nucleic acid sequence different from said previous substrate nucleic acid sequence, and encoding the same protein or peptide; and

5 (l) repeating steps (c) through (j) above with said subsequent substrate nucleic acid sequence.

10 10. A method according to claim 8 or 9, wherein said step (l) further comprises repeating steps (c) through (k) until a nucleic acid sequence having sufficient efficiency of translation in said host is identified.

11. A method according to claim 1, further comprising the steps of :
detecting a phase shift in said three base periodic cycle; and
determining the presence of a frame shift in said substrate nucleic acid
15 sequence from said phase shift, so that said substrate nucleic acid sequence remains a candidate for efficient translation in said host in the presence of said phase shift.

12. A method according to claim 1, wherein said host is a prokaryotic species.

20 13. A method according to claim 1, wherein said host is a eukaryotic species.

14. A method according to claim 1, further comprising the step of:
predicting the potential for proper translation of said substrate nucleic acid
sequence in said host using the said already determined binding strengths of said
25 ribosomal nucleic acid sequence to said substrate nucleic acid sequence.

15. A method according to claim 1, wherein said step (g) of generating a binding strength pattern comprises one or more calculating step(s), said calculating step(s) selected from a group comprising (i) calculating a summation of all said
30 binding strengths, (ii) calculating a summation of said first, second, or third binding strengths, (iii) calculating an integral of all said binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding

strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said binding strengths or of said first, second, or third binding strengths.

5

16. A method according to claim 1, wherein said step (h) of detecting the presence or absence of a three-base binding strength periodic cycle comprises one or more calculating step(s), said calculating step(s) selected from a group comprising (i) calculating a summation of all said binding strengths, (ii) calculating a summation of
10 said first, second, or third binding strengths, (iii) calculating an integral of all said binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said
15 binding strengths or of said first, second, or third binding strengths.

17. A method according to claim 1, wherein said step (f) of successively repeating steps (c) through (e) further comprising the step of:

calculating a cumulative energy for each said first, second, or third alignment
20 of interest.

18. A method according to claim 1, further comprising the step of:

calculating a cumulative energy differential for a portion of said substrate nucleic acid sequence.
25

19. A method according to claim 1, further comprising the step of:

calculating a power spectrum magnitude of said binding strengths of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence for a portion of said substrate nucleic acid sequence.
30

20. A method according to claim 1, further comprising the step of:

calculating a mean binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence for a portion of said substrate nucleic acid sequence.

5 21. A system for screening a nucleic acid sequence for efficient translation in a predetermined host, comprising:

 (a) means for providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said
10 host;

 (b) means for providing a substrate nucleic acid sequence;

 (c) means for determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest in the substrate;

15 (d) means for determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second substrate alignment of interest is one base downstream from said first alignment of interest;

 (e) means for determining a third binding strength of said ribosomal nucleic
20 acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third substrate alignment of interest is two bases downstream from said first alignment of interest;

 (f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step
25 (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

 (g) means generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

30 (h) means for detecting the presence or absence of a three-base binding strength periodic cycle and phase through said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength

cycle and correct phase through said substrate nucleic acid sequence indicating said substrate nucleic acid sequence is a candidate for efficient translation in said host.

22. A computer program product for screening a nucleic acid sequence for
5 efficient translation in a predetermined host, said computer program product comprising a computer usable storage medium having computer readable program code means embodied in the medium, the computer readable program code means comprising:

(a) means for providing a ribosomal nucleic acid sequence, said ribosomal
10 nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

(b) means for providing a substrate nucleic acid sequence;

(c) means for determining a first binding strength of said ribosomal nucleic
15 acid sequence to said substrate nucleic acid sequence at a first alignment of interest in the substrate;

(d) means for determining a second binding strength of said ribosomal nucleic
acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second substrate alignment of interest is one base downstream
20 from said first alignment of interest;

(e) means for determining a third binding strength of said ribosomal nucleic
acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third substrate alignment of interest is two bases downstream from said first alignment of interest;

25 (f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

30 (g) means for generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

(h) means for detecting the presence or absence of a three-base binding strength periodic cycle and phase through said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle and correct phase through said substrate nucleic acid sequence indicating said substrate nucleic acid sequence is a candidate for efficient translation in said host.

23. A computer-based method for screening a nucleic acid sequence for the presence of at least one coding sequence therein, said method comprising the steps of:

(a) providing a substrate nucleic acid sequence from a predetermined host;
10 (b) providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

(c) determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest;

15 (d) determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second alignment of interest is one base downstream from said first alignment of interest;

(e) determining a third binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third alignment of interest is two bases downstream from said first alignment of interest;

20 (f) successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

(g) generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

30 (h) detecting the presence or absence of a three-base binding strength periodic cycle and phase in at least one portion of the said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding

strength cycle through said substrate nucleic acid sequence indicating said at least one portion of said nucleic acid sequence is a coding portion in said host.

24. A method according to claim 23, further comprising the step of:

- 5 (i) determining a start region and a stop region of said coding portion within said substrate nucleic acid sequence.

25. A method according to claim 23, wherein said detecting further comprises the step of determining the strength of said periodic signal.

10

26. A method according to claim 25, further comprising the step of:

- (i) generating a quantitative indicator of translation efficiency from the strength of said periodic signal.

15

27. A method according to claim 23, further comprising the steps of :

detecting a phase shift in said three base periodic cycle; and

determining the presence of a frame shift in said substrate nucleic acid sequence from said phase shift, so that said substrate nucleic acid sequence remains a candidate for efficient translation in said host in the presence of said phase shift.

20

28. A method according to claim 23, wherein said host is a prokaryotic species.

29. A method according to claim 23, wherein said host is a eukaryotic species.

25

30. A method according to claim 23, further comprising the step of:

predicting the potential for proper translation of said substrate nucleic acid sequence in said host using the said already determined binding strengths of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence.

30

31. A method according to claim 23, wherein said step (g) of generating a binding strength pattern comprises one or more calculating step(s), said calculating

step(s) selected from a group comprising (i) calculating a summation of all said binding strengths, (ii) calculating a summation of said first, second, or third binding strengths, (iii) calculating an integral of all said binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said binding strengths or of said first, second, or third binding strengths.

10 32. A method according to claim 23, wherein said step (h) of detecting the presence or absence of a three-base binding strength periodic cycle comprises one or more calculating step(s), said calculating step(s) selected from a group comprising (i) calculating a summation of all said binding strengths, (ii) calculating a summation of said first, second, or third binding strengths, (iii) calculating an integral of all said binding strengths or of said first, second, or third binding strengths, (iv) calculating a partial integral of all said binding strengths or of said first, second, or third binding strengths, (v) calculating a running average of all said binding strengths or of said first, second, or third binding strengths, and (vi) calculating transforms of all said binding strengths or of said first, second, or third binding strengths.

20 33. A method according to claim 23, wherein said step (f) of successively repeating steps (c) through (e) further comprising the step of:

calculating a cumulative energy for each said first, second, or third alignment of interest.

25 34. A method according to claim 23, further comprising the step of:
calculating a cumulative energy differential for a portion of said substrate nucleic acid sequence.

30 35. A method according to claim 23, further comprising the step of:

calculating a power spectrum magnitude of said binding strengths of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence for a portion of said substrate nucleic acid sequence.

5 36. A system for screening a nucleic acid sequence for the presence of at least one coding sequence therein, comprising:

 (a) means for providing a substrate nucleic acid sequence from a predetermined host;

 (b) means for providing a ribosomal nucleic acid sequence, said ribosomal
10 nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

 (c) means for determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest;

15 (d) means for determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second alignment of interest is one base downstream from said first alignment of interest;

 (e) means for determining a third binding strength of said ribosomal nucleic
20 acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third alignment of interest is two bases downstream from said first alignment of interest;

 (f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step
25 (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

 (g) means for generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e);
30 and

 (h) means for detecting the presence or absence of a three-base binding strength periodic cycle and phase in at least one portion of the said substrate nucleic

acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle through said substrate nucleic acid sequence indicating said at least one portion of said nucleic acid sequence is a coding portion in said host.

5

37. A computer program product for screening a nucleic acid sequence for the presence of at least one coding sequence therein, said computer program product comprising a computer usable storage medium having computer readable program code means embodied in the medium, the computer readable program code means comprising:

10

(a) means for providing a substrate nucleic acid sequence from a predetermined host;

(b) means for providing a ribosomal nucleic acid sequence, said ribosomal nucleic acid sequence comprising a continuous segment of from 6 to 15 bases of the first 15 bases closest to the 3' end of the 16S or 18S rRNA ribosomal subunit of said host;

15

(c) means for determining a first binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a first alignment of interest;

(d) means for determining a second binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a second alignment of interest; wherein said second alignment of interest is one base downstream from said first alignment of interest;

20

(e) means for determining a third binding strength of said ribosomal nucleic acid sequence to said substrate nucleic acid sequence at a third alignment of interest, wherein said third alignment of interest is two bases downstream from said first alignment of interest;

25

(f) means for successively repeating steps (c) through (e) along said substrate nucleic acid sequence, wherein the first alignment of interest in each successive step (c) is three bases downstream of the first alignment of interest in the immediately preceding step (c), until a binding strength is determined at every alignment of said substrate nucleic acid sequence;

30

(g) means for generating a binding strength pattern from said first through third binding strengths determined in said successively repeated steps (c) through (e); and

(h) means for detecting the presence or absence of a three-base binding strength periodic cycle and phase in at least one portion of the said substrate nucleic acid sequence from said binding strength pattern, the presence of said three base periodic binding strength cycle through said substrate nucleic acid sequence indicating said at least one portion of said nucleic acid sequence is a coding portion in said host.

10

10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532